

# Evaluating Meta-analyses in the General Surgical Literature

## *A Critical Appraisal*

*Elijah Dixon, MD, MSc, FRCSC,\* Morad Hameed, MD, MSc, FRCSC,\*†  
Francis Sutherland, MD FRCSC,\* Deborah J. Cook, MD, MSc, FRCPC,§ and  
Christopher Doig, MD, MSc, FRCPC†‡*

**Objective:** To assess the methodologic quality of meta-analyses of general surgery topics published in peer-reviewed journals.

**Summary Background Data:** Systematic reviews and meta-analysis are used to seek, summarize, and interpret primary studies on a given topic. Accordingly, systematic reviews and meta-analyses of high-quality primary studies may be the highest level of evidence for issues of prevention and treatment in evidence-based medicine. However, not all published meta-analyses are rigorously performed.

**Methods:** We searched MEDLINE (from January 1, 1997, to September 1, 2002) and reference lists and solicited general surgery specialists to identify relevant meta-analyses. Inclusion criteria were use of meta-analytic methods to pool the results of primary studies in general surgery on issues of diagnosis, causation, prognosis, or treatment. Our search strategies identified 487 potentially relevant articles. After excluding articles based on a priori criteria, 51 meta-analyses fulfilled eligibility criteria. In duplicate and independently, 2 reviewers assessed the quality of these meta-analyses using a 10-item index called the Overview Quality Assessment Questionnaire.

**Results:** Overall concordance between 2 independent reviewers was good (interobserver agreement 81%, and a  $\kappa$  of 0.62 (95% CI 0.55–0.69). Of 51 relevant articles, 38 were published in surgical journals. Most studies had major methodologic flaws (median score of 3.3, scale of 1–7). Factors associated with low overall scientific quality included the absence of any prior meta-analyses publications by authors and meta-analyses produced by surgical department members without external collaboration.

**Conclusions:** This critical appraisal of meta-analyses published in the general surgery literature demonstrates frequent methodologic flaws. The quality of these reports limits the validity of the findings and the inferences that can be made about the primary studies reviewed. To improve the quality of future meta-analyses, we

recommend following guidelines for the optimal conduct and reporting of meta-analyses in general surgery.

(*Ann Surg* 2005;241: 450–459)

The profusion of publications in scientific and biomedical journals makes it difficult for busy clinicians, educators, and investigators to keep abreast of new developments. For this reason, systematic reviews and meta-analysis are used to seek, summarize, and interpret primary studies on a given topic. These publications have been used to inform clinical practice, to aid teaching, direct health policy, guide future research, and to serve as a foundation for practice guidelines.<sup>1</sup> Accordingly, systematic reviews and meta-analyses of high-quality primary studies are the highest level of evidence for issues of prevention and treatment in evidence-based medicine.<sup>2</sup> Indeed, the Oxford Centre for Evidence Based Medicine ([www.indigojazz.uk/cebmd](http://www.indigojazz.uk/cebmd)) ranks systematic reviews/meta-analyses as level 1a evidence.

However, not all published meta-analyses are rigorously performed.<sup>3–13</sup> Moreover, the results of meta-analyses have been criticized because sometimes they differ from the results of subsequent large randomized trials.<sup>5,14–17</sup> The discordance between meta-analyses and subsequent randomized trials may be in part due to these shortcomings in meta-analysis methodology.<sup>3,5,6,8,11,18</sup>

The Overview Quality Assessment Questionnaire (OQAQ) was developed as a tool for the critical appraisal of the quality of meta-analyses.<sup>19</sup> Its operating characteristics have been validated; including interrater reliability, face validity, and construct validity as measured against 7 a priori hypotheses dealing with how the instrument should perform if adequately measuring scientific quality of systematic reviews.<sup>19–21</sup> The OQAQ has been used to assess publications in both the emergency medicine and anesthesia literature. Using the OQAQ, our objective was to assess the quality of meta-analyses published in the general surgery literature. Our primary goals were to assess general surgery topics reviewed

From the \*Departments of Surgery, †Critical Care, and ‡Community Health, Faculty of Medicine, University of Calgary, Calgary, Alberta, Canada; and the §Department of Medicine and Clinical Epidemiology, McMaster University, Hamilton, Ontario, Canada.

Reprints: Christopher James Doig, MD, MSc, University of Calgary, Foothills Hospital, Division of Critical Care, 1403–29<sup>th</sup> Street NW, Calgary, Alberta, Canada T2N 2T9. E-mail: [Chip.Doig@CalgaryHealthRegion.ca](mailto:Chip.Doig@CalgaryHealthRegion.ca).

Copyright © 2005 by Lippincott Williams & Wilkins

ISSN: 0003-4932/05/24103-0450

DOI: 10.1097/01.sla.0000154258.30305.df

using the technique of meta-analysis, to evaluate the rigor of several specific steps in the conduct of meta-analyses, and to identify areas of weakness to target for improvement in future reviews.

## MATERIALS AND METHODS

### Inclusion Criteria and Search Strategy

We defined a “meta-analysis” as a systematic review which includes “a statistical analysis of the results from independent studies, which generally aims to produce a single estimate of effect.”<sup>22</sup> A literature review was performed to identify all meta-analyses of general surgery topics published in peer-reviewed paper-based journals. We performed the literature search using the PubMed (MEDLINE) search engine in November 2002 for the period January 1, 1997, to September 1, 2002 (search strategy outlined in Fig. 1). The search term used was *surgery* (276,268 articles) limited using the “limits” to the following: meta-analysis publication types (528 articles), human only studies (526 articles), and English language (487 articles). All abstracts were then reviewed by 1 investigator to identify studies for inclusion. Inclusion criteria were (1) use of meta-analytic methodology to pool the results of primary studies; (2) issues of diagnosis, causation, prognosis, or treatment; and (3) focus on conditions relating directly to general surgery practice. Publications were excluded if they (1) were neither meta-analyses nor systematic reviews ( $N = 145$ ); (2) because they were systematic reviews but not meta-analyses (eg, the primary study results were not pooled statistically) ( $N = 114$ ); (3) they did not address general surgery topics ( $N = 165$ ) (22 neurosurgery, 22 chemo/radiotherapy, 21 cardiothoracic, 19 orthopedics, 15 anesthesia/analgesia, 10 otolaryngology, 9 obstetrics and gynecology, 8 vascular surgery, 8 gastroenterology, 6 pediatric surgery, 5 ophthalmology, 5 dentistry, 4 plastic surgery, 4 urology, 3 critical care, 2 transplantation, and 2 miscellaneous); and (4) they were Cochrane Reviews ( $N = 12$ ), since prior research has shown these to be of high quality.<sup>23</sup> We did not consider duplicate publications, unpublished work, abstracts, and conference proceedings. We identified 51 relevant meta-analyses.

### Quality Appraisal of Articles and Overview Methods

All meta-analyses were critically appraised independently in duplicate (ED, CD), using the OQAQ. This checklist includes 9 items (scored as done, partially done/cannot tell, or not done), and a 10th item requiring a summary evaluation.<sup>19–21</sup> The OQAQ has been psychometrically tested and found to be valid and reliable (interrater reliability, face validity, and construct validity).<sup>19–21</sup> When scoring items 1 through 9, we scored “partially” if methods were reported incompletely, or “cannot tell” if methods were not reported at

all. These items were scored as “yes” or “no” only when the criterion was explicitly met or not met.<sup>19</sup> The 10th item is an overall assessment of scientific quality on a scale of 1 to 7 (1 indicating extensive flaws with major risk of bias to 7 indicating minimal flaws with minor risk of bias). This score is based on the results of the preceding 9 items, and we followed the published recommendations for scoring. If a meta-analysis scored “cannot tell” on 1 or more of the 9 core items, we considered it to have minor flaws at best and it received a score of 4 or lower. If the meta-analysis scored a “no” on question 2, 4, 6, or 8, we considered it to have, at a minimum, major flaws, and it received a score of 3 or less.<sup>19</sup> Final scores were obtained by consensus of the 2 reviewers. Overall concordance was good, with an interobserver agreement rate of 81% and a  $\kappa$  coefficient of 0.62 (95% CI 0.55–0.69) prior to consensus. When consensus between the 2 reviewers could not be reached, a third reviewer was used to adjudicate the final score (as occurred for 2 meta-analyses) on points of disagreement (MH, FS).

To identify publication evidence of expertise in meta-analytic methods, we also performed a PubMed search of all listed authors of the included meta-analyses using the PubMed limit “meta-analysis” to identify the number of prior meta-analyses published. Data regarding the number of patients and studies included in the meta-analysis were abstracted when possible. We recorded the department producing the meta-analysis. Finally, the impact factors specific to the year of publication were obtained for the host journal (Institute for Scientific Information).

### Analysis

Using the methods of Spearman, correlation coefficients are reported for associations examined between covariates and the summary score. Summary effect sizes were extracted from each meta-analysis. When available, we used odds ratios (ORs) for mortality. If raw data were reported, we calculated ORs. Otherwise, we abstracted relative risks as the main summary statistic. All metrics were converted such that values (OR or RR) greater than 1 favored the experimental treatment over the control. Statistical analysis was done using SAS software.

## RESULTS

In Table 1, we present the meta-analyses and their characteristics. In Table 2, we present the component scores (9 core questions) of all 51 articles reviewed. Sixty-five percent of all meta-analyses used comprehensive search methods, and 67% clearly reported their search strategies. Inclusion criteria were described in 70% of the studies reviewed, although 10% did not describe inclusion criteria. Between 14% and 43% of the meta-analyses did not adequately avoid bias in the selection of studies included in the meta-analysis. These publications were weakest with regard

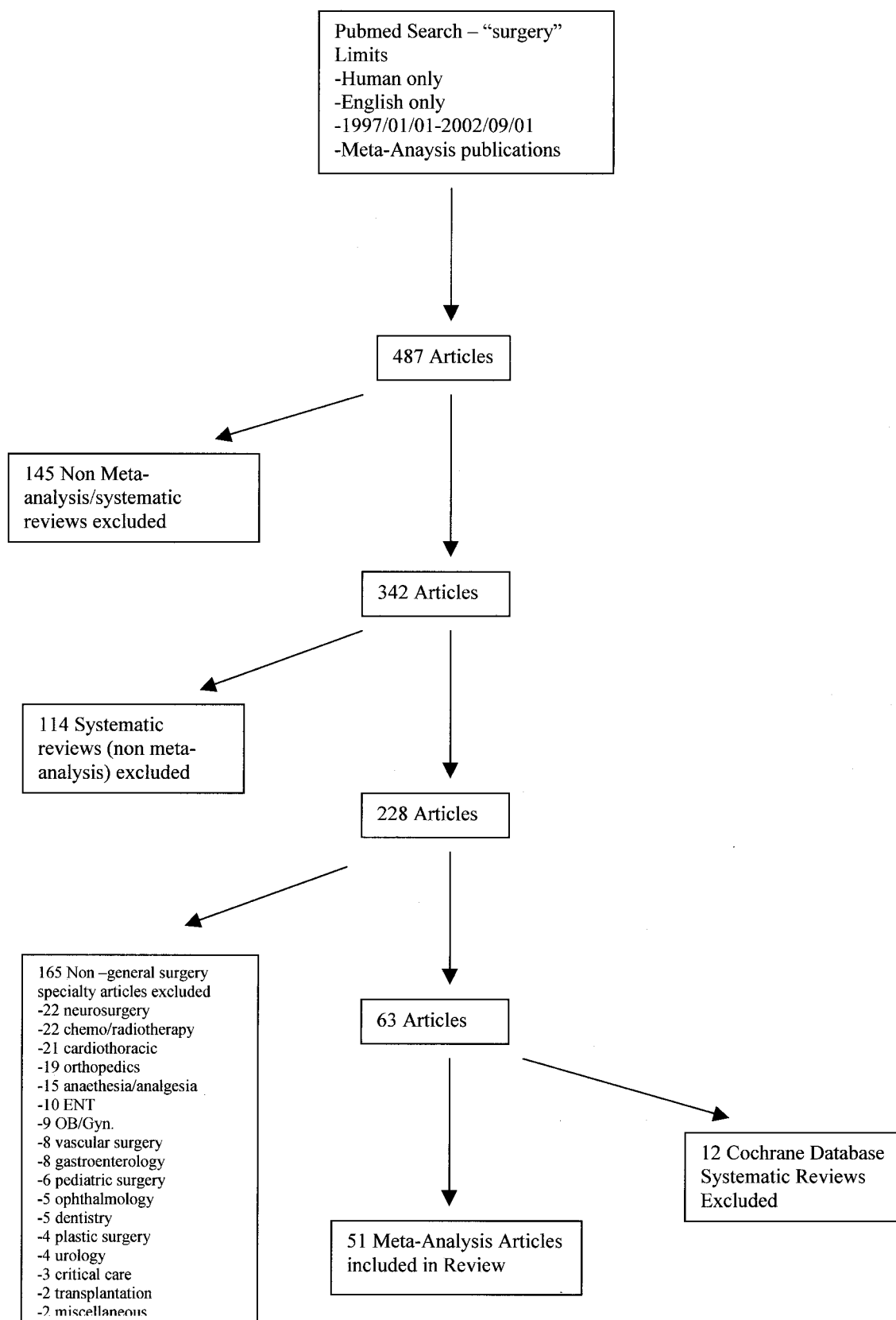


FIGURE 1. Literature search strategy.

TABLE 1. References Included in Review

Reference	Year	Topic	Total # Patients	# Of studies in MA	Journal Impact Factor	OSQ	Department/s Producing Publication	Outcome	Summary Effect: Odds Ratio (Relative Risk)*
27	1997	Miscellaneous	547	6	0.936	5	Other	Time to first flatus	
28	1997	Surgical technique	682	6	0.502	5	Public Health	Quality of life	
29	1997	Anticoagulation	16,583	36	2.287	3	Surgery and Public Health	Incidence of DVT, wound hematoma	1.14
30	1997	Oncology	17,815		2.287	1	Surgery	Survival	
31	1997	Surgical technique	4460	7		2	Other	Survival	1.11
32	1997	Oncology		14	2.344	2	Surgery	Incidence of colorectal cancers	0.92
33	1997	Anticoagulation		33	1	2	Public Health	Incidence DVT, bleeding	0.93*
34	1998	Surgical technique		45	0.827	3			
35	1998	Surgical technique	2256	13	2.138	7	Surgery	Mortality/morbidity/cancer recurrence	1.01
36	1998	Oncology	2532	8	2.381	7	Other	Mortality/infection/recurrence	1.0
37	1998	Surgical technique	452	3	2.138	1	Surgery	Wound infection/anastomotic leak	
38	1998	Oncology	2005	5	2.138	2	Surgery	Survival	
39	1998	Anticoagulation		57	0.994	1	Surgery	Incidence DVT/PE	3.33
40	1998	Wound closure	12,249	25	1.874	2	Surgery	Incidence of infection/hernia/dehiscence	
41	1999	Oncology			3.632	2	Surgery	Recurrence rates	
42	1999	Miscellaneous			3.072	1	Other	Rebleed rate	3.8
43	1999	Laparoscopy	2471	14	2.24	4	Surgery and Public Health	OR time, pain, return to normal activity, recurrence	0
44	1999	Laparoscopy	1373	11	.	5	Other	OR time, pain, return to normal activity, wound infection	
45	1999	Oncology	1009	11	5.647	3	Surgery	Mortality, morbidity, LOS	2.13
46	1999	Surgical technique	2727	17	1.926	2	Surgery	Incontinence, persistence of fissure	1.16*
47	1999	Surgical technique/Oncology	2936	6	2.427	2	Surgery	Survival	
48	1999	Miscellaneous		4		2	Surgery	Fistula rate	
49	1999	Laparoscopy	1383	12	0.527	7	Surgery	OR time, LOS, morbidity, return to normal activity, readmission	2.5
50	1999	Surgical technique	414	4	5.647	6	Other	Morbidity	1.4
51	1999	Miscellaneous	389	7	0.824	2	Other	Incidence of adhesions	2.86
52	2000	Miscellaneous		420	2.116	3	Surgery		
53	2000	Wound closure	5145	13	5.987	7	Surgery	Incisional hernia, wound dehiscence, infection, wound pain, suture sinus	1.47
54	2000	Oncology	3486	19	2.456	4	Other	Ability to predict presence of cancer	5.52

(Continued)

TABLE 1. (Continued)

Reference	Year	Topic	Total # Patients	# Of studies in MA	Journal Impact Factor	OSQ	Department/s Producing Publication	Outcome	Summary Effect: Odds Ratio (Relative Risk)*
55	2000	Surgical technique		10	1.464	4	Public Health	Weight reduction, OR morbidity	
56	2000	Surgical Technique	7241	35	1.674	2	Surgery	Postoperative thyroid function, morbidity	
57	2000	Surgical technique	2532	27	2.456	1	Surgery	Wound infection	
58	2000	Miscellaneous		51	2.763	4	Other	Worbidity	1.59
59	2000	Oncology/Laparoscopy			2.629	1	Surgery	Incidence of wound recurrence	
60	2000	Surgical technique	303	3	2.13	7	Surgery	Treatment failures, morbidity, 30 d mortality, number of retreatments	1.92
61	2000	Surgical technique	506	15	1.69	2	Surgery	Reoperation rate	
62	2000	Surgical technique	865	5	3.489	3	Other	Sensitivity	
63	2001	Oncology		52	5.647	3	Other		
64	2001	Surgical technique	4941	50	2.022	3	Surgery	Mortality, morbidity, survival	1.1
65	2001	Miscellaneous	837	11	5.143	3	Other	Mortality, morbidity, LOS	1.39*
66	2001	Anticoagulation	5520	8	2.732	3	Other	Incidence DVT/PE	3.57*
67	2001	Wound closure	5718	15	0.992	2	Surgery	Wound dehiscence, pain, infection, hernia, suture sinus	1.93*
68	2001	Surgical technique	342	6	1.721	5	Surgery	Mortality, morbidity	1.92
69	2001	Surgical technique	467	5	.	7	Surgery	Mortality, morbidity	2.70
20	2002	Surgical technique	11,174	58	6.674	2	Public Health	Recurrence, pain	2.33
70	2002	Miscellaneous	624	12	3.464	7	Public Health	Transfusion rates	3.85
71	2002	Oncology	1523	3	2.792	3	Public Health	Mortality	1.16
72	2002	Antibiotics	2065	13	0.503	2	Surgery	Surgical site infection	3.45*
73	2002	Surgical technique	327,523	40	2.374	1	Surgery	Incidence, type, timing of cbd injury diagnosis	
74	2002	Oncology	1342	5	6.629	3	Surgery	All cause 5-y mortality, recurrence rates	1.37
75	2002	Surgical technique	3155	23	6.674	7	Surgery	In-hospital death, morbidity, LOS	0.84
76	2002	Laparoscopy	4688	27	2.131	2	Surgery	Return to work, morbidity, cost	

OSQ indicates Overall Scientific Quality, Question 10 of OQAQ.

\*Relative risk.

to reporting explicit inclusion criteria, the description of validity criteria used to select studies to include in the meta-analysis, and the appropriateness of these criteria. Seventy-one percent of the articles had methodologic flaws in the description of validity criteria, while 39% to 70% lacked or did not use appropriate validity criteria. Items 7 and 8 of the OQAQ focus on the reporting and appropriateness of the way

in which the studies were combined. Sixty-seven percent of the articles reported how the results of the individual studies were combined, and 65% of the studies combined the results appropriately. Question 9 assesses whether or not the conclusions are supported by the results, in whole or in part. These meta-analyses articles scored the highest on this item; 78% had conclusions supported by the data reported.



**TABLE 2.** Summary of Questions 1–9: Core Index Questions

Ten Index Questions	No. (%)	Partially/Cannot Tell (%)	Yes (%)
1. Were the search methods used to find evidence stated?	4	29	67
2. Was the search for evidence reasonably comprehensive?	6	29	65
3. Were the criteria used for deciding which studies to include in the review reported?	10	20	70
4. Was bias in the selection of studies avoided?	14	29	57
5. Were the criteria used for assessing the validity of the included studies reported?	61	10	29
6. Was the validity of all studies referred to in the text assessed using appropriate criteria?	39	31	29
7. Were the methods used to combine the findings of the studies reported?	20	14	67
8. Were the findings of the relevant studies combined appropriately relative to the primary question the overview addresses?	16	20	65
9. Were the conclusions made by the author(s) supported by the data and/or analysis reported in the overview?	2	20	78

Cumulative responses to question 10 (overall scientific quality, OSQ) include: median score of 3 and a mean score of 3.33. The scores for these 51 meta-analyses were distributed as follows: 7 (14% for score of 1), 16 (31%, score of 2), 11 (22%, score of 3), 4 (8%, score of 4), 4 (8%, score of 5), 1 (2%, score of 6), and 8 (16%, score of 7).

Table 3 shows that, on average, 27 (3–420) studies and 11,853 (303–327,523) patients were included in each meta-analysis. The mean impact factor for the journal of publication was 2.684 (0.502–6.674). PubMed was searched using each author's name included in the article, and the total per publication was summed (eg, if a publication had 5 authors and 4 authors had previously published 1 meta-analysis each and the fifth had published 2, then the total for that paper was 6 prior meta-analysis publications). When authors on a pub-

lication had coauthored 1 prior meta-analysis together, it was only scored as 1. A summary “density” score was then calculated by dividing the total number of prior meta-analyses published by the number of authors on the paper. The mean “density” score for all 51 articles is 1.48 (0.2–8.3). The cumulative authors per publication had 4.78 (0–25) other meta-analysis publications. The following factors were not significantly correlated with OSQ: the impact factor of the journal in which the meta-analysis was published, the number of patients and studies analyzed in the meta-analysis, and the summary measure of effect (OR or RR) (Table 3). A significant positive correlation was detected between the OSQ and the number of previous meta-analyses published by authors of the index meta-analysis. Analysis by the number of other meta-analyses published by all the authors of a given paper is contained in Table 4. The majority of meta-analysis authors (40 of 51) had published meta-analyses previously. Those papers produced by authors with prior meta-analysis published have a significantly higher mean OSQ score compared with those by first-time authors (3.55 versus 2.55).

Various subgroup analyses are outlined in Table 4. The number of meta-analyses published per year is shown, with a range of 7 to 11. There is no significant difference in mean OSQ by year of publication. We list the number of publications per journal in decreasing order of mean OSQ. The *Annals of Surgery* had the highest number of publications by a single journal (6 meta-analyses). Notably, this journal also had the highest mean OSQ (4.67). Journals publishing at least 2 meta-analyses were analyzed separately. Those surgical journals publishing 1 meta-analysis were grouped together (mean OSQ, 3.250). All nonsurgical journals were grouped together (“all other publications,” mean OSQ, 3.308). We identified a substantial range in quality of these meta-analyses, stratified by journal of publication. We also categorized articles by topic; the mean OSQ listed by topic, from greatest

**TABLE 3.** Miscellaneous Characteristics of Publications and Correlation Between Factors

Meta-analysis Characteristics	Mean (Range)
Impact factor of journals meta-analysis published in	2.684 (0.502–6.674)
# Of studies included in meta-analysis	27 (3–420)
# Of patients included in meta-analysis	11,853 (303–327,523)
# Of other meta-analysis publications by authors of each paper	4.78 (1–25)
Factors	Correlation coefficients
Odds ratio (Effect) × summary score	−0.12 (NS)
Impact factor × summary score (question 10)	0.19 (NS)
# Of studies in MA × summary score	−0.12 (NS)
# Of other MA publications × summary score	0.44 (0.0012)
# Of patients in MA × summary score	−0.23 (NS)

**TABLE 4.** Analysis of Factors Determining Overall Quality of Publications

Year	# Of Meta-analysis	Mean Overall Score (by Year)
1997	8	3.125
1998	7	3.286
1999	10	3.100
2000	11	3.455
2001	7	3.714
2002	8	3.375
Overall mean score by journal of publication		
<i>Annals of Surgery</i>	6	4.670
<i>Canadian Journal of Surgery</i>	2	4.500
<i>British Journal of Surgery</i>	5	4.200
All other publications	13	3.308
All other surgical journals	8	3.250
<i>American Journal of Surgery</i>	4	3.000
<i>Diseases of the Colon and Rectum</i>	5	2.800
<i>Surgery</i>	2	2.500
<i>Surgical Endoscopy</i>	2	2.500
<i>Archives of Surgery</i>	2	2.000
<i>American Surgeon</i>	2	1.500
Overall mean score by topic of publication		
Laparoscopy	4	4.500
Surgical technique	20	3.600
Wound closure	3	3.667
Miscellaneous	8	3.375
Oncology	11	2.818
Anticoagulation	4	2.250
Antibiotic therapy	1	2.000
Overall mean score by department producing publication		
Public Health	6	3.833
Other	13	3.692
Surgery and Public Health	2	3.500
Surgery	29	3.069
Overall mean score by number of meta-analysis publication produced by author		
0	9	2.555
≥1	40	3.550

to least: laparoscopy, surgical technique, wound closure, miscellaneous, oncology, anticoagulation, and antibiotic therapy. The articles were also categorized according to the department(s) that produced the publication. The number of publications, along with the mean OSQ by department of

publication, is listed in Table 4. By mean OSQ, from greatest to least, the order of department(s) of publication is Public Health/Epidemiology, Other, Surgery and Public Health/Epidemiology, and Surgery.

## DISCUSSION

To the best of our knowledge this is the first review of meta-analyses published on general surgical topics. Despite recognition of the importance of strategies to limit bias in the assembly and analysis of primary studies included in meta-analyses,<sup>19,24</sup> our review demonstrates that this knowledge has not been well implemented in the general surgical literature. When conducted rigorously, systematic reviews may usefully guide practice, teaching, research, and health policy.<sup>2</sup> However, when bias in the primary studies and/or the review methods is not minimized, the results and conclusions of meta-analyses may not be valid.<sup>5,25</sup> Dissemination of meta-analyses with significant methodologic flaws may lead readers to abandon systematic reviews altogether or misinterpret or misuse them.

Overall, we found that the quality of these surgical meta-analyses is low. We have identified some factors associated with both high and low overall quality scores. Two factors that correlate with the overall score of the articles relate to the department producing the publication, as well as the meta-analysis publication experience of the authors. Studies produced by groups in which at least 1 author was a member of a department of public health or epidemiology unit had the highest scores, whereas meta-analyses produced by authors all of whom were members of a surgical department had the lowest. As well, our data demonstrate that meta-analyses by authors with at least 1 previously published meta-analysis have higher overall mean scores than other published meta-analyses. We suggest that meta-analyses should be authored by a group of individuals with both clinical expertise and methodologic expertise. In those circumstances where none of the authors have prior methodologic expertise or experience, an expert should be consulted and the QUOROM guidelines should be followed.

We found that the range of mean OSQ by journal of publication is wide, 4.670 to 1.500. Although the mean summary score did not significantly correlate with the impact factor of the journal, 2 of the 3 journals with the highest mean scores are both high impact surgical journals (*Annals of Surgery* and the *British Journal of Surgery* [BJS]). The BJS instructions to authors contain a referral to the use of the QUORUM criteria<sup>9</sup> for those who are going to submit a systematic review/meta-analysis. To some degree, it is the responsibility of peer reviewers and editors to ensure that careful steps are taken to ensure a valid meta-analysis and transparent reporting of systematic review methods. Editorial interest in these issues may partially explain the high-quality meta-analyses in certain journals.

The methodologic areas of weakness we identified include validity assessment, selection bias, reporting of search strategies, and pooling of data. Kelly et al<sup>26</sup> have outlined strategies that can be used to both assess and improve the scientific quality of meta-analyses/systematic reviews; these are outlined below. Validity assessment can be improved by the use of a validated scoring system to grade the quality of studies included in the meta-analysis. Similarly, minimizing selection and ascertainment bias can be accomplished by the use of multiple assessors, blinding of assessors, adjudication, and measurement reporting. Reporting of search strategies along with comprehensive search strategies should decrease bias. Search strategies include MEDLINE, EMBASE, and other bibliographic databases, seeking unpublished literature, contact with authors, hand searching, and inclusion of non-English language. The use of methodologic guidelines such as those incorporated in the QUOROM statement should help to minimize bias and poor reporting of meta-analyses.

By comparison, this review of the surgical literature fall somewhat short of the anesthesia literature; 41.5% of the systematic reviews in a recent anesthesia report had minor or minimal flaws.<sup>1</sup> In this review, the mean OSQ for the anesthesia literature was 4.3. Similarly, the emergency medicine literature was reviewed using the OQAQ; only 13% had minor flaws.<sup>26</sup> The mean overall score was 2.7, which is significantly lower than the mean score of 3.33 for our present review. Thus, although we found significant shortcomings in meta-analyses in the field of general surgery, the surgical literature is keeping with the quality of published meta-analyses in other fields of medicine. Despite our findings in the field of general surgery, there are some encouraging trends. Figure 2 demonstrates the mean OSQ and

impact factor by year of publication. Over time, meta-analyses are being published in journals with impact factors of increasing value. Possible explanations for this change include dissemination and acceptance of the value of meta-analytic methods by both researchers and peer reviewers and an increase in the impact factors of surgical journals. The trend in OSQ over time is less robust and may demonstrate a slow improvement in overall quality.

An important feature of our analysis reinforces the potential relationship between the overall quality of published meta-analyses and both the direction and magnitude of treatment effect estimates. Estimates of effect (OR and relative risks) demonstrate a weak correlation between OSQ and the magnitude of effect, which is not statistically significant (Table 3). We found that as the quality of the meta-analysis decreases, the magnitude of effect increased (correlation coefficient =  $-0.12$ ). This “overestimation” of “effect” has been demonstrated in prior studies examining both randomized trials and meta-analyses,<sup>10,11</sup> with overestimates of effect being as high as 41%. Studies of low quality therefore have results that may not be valid and more importantly demonstrate systematic error, or a bias favoring the “experimental treatment.”

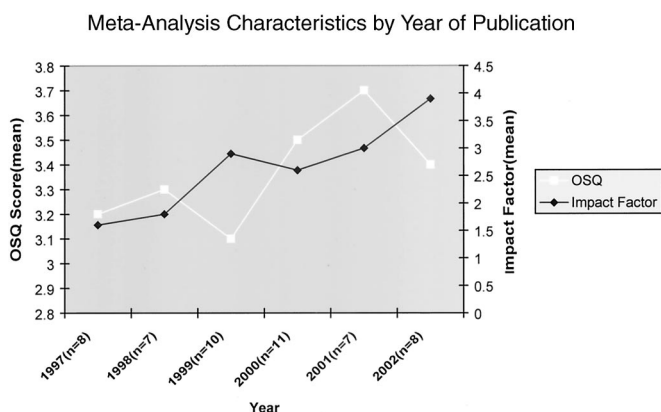
Overall, the scientific quality of meta-analyses published on topics pertaining to general surgery is low, and the majority have methodologic flaws. This may impair the validity of these publications and thus limit their use for clinical, educational, research, and policy purposes. In the future, more attention to rigorous systematic review methods by authors, constructively critical suggestions by peer reviewers, and attention to the QUOROM statement recommendations by editors should lead to improvement in these important publications in the field of surgery.

## ACKNOWLEDGMENTS

Dr. D. Cook is a Research Chair of the Canadian Institutes for Health Research.

## REFERENCES

1. Choi PT, Halpern SH, Malik N, et al. Examining the evidence in anesthesia literature: a critical appraisal of systematic reviews. *Anesth Analg*. 2001;92:700–709.
2. Guyatt GH, Sackett DL, Sinclair JC, et al. Users' guides to the medical literature, IX: a method for grading health care recommendations: Evidence-Based Medicine Working Group. *JAMA*. 1995;274:1800–1804.
3. Balk EM, Bonis PA, Moskowitz H, et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. *JAMA*. 2002;287:2973–2982.
4. Christensen E. Quality of reporting of meta-analyses: the QUOROM statement: will it help? *J Hepatol*. 2001;34:342–345.
5. Flather MD, Farkouh ME, Pogue JM, et al. Strengths and limitations of meta-analysis: larger studies may be more reliable. *Control Clin Trials*. 1997;18:568–579.
6. Ioannidis JP, Cappelleri JC, Lau J. Issues in comparisons between meta-analyses and large trials. *JAMA*. 1998;279:1089–1093.



**FIGURE 2.** Characteristics of meta-analysis by year of publication. White line represents mean OSQ score by year (left vertical axis). Black line represents mean impact factors by year (right vertical axis). Horizontal axis represents year of publication, with n equal to number of articles published for a given year.



7. Jadad AR, Cook DJ, Jones A, et al. Methodology and reports of systematic reviews and meta-analyses: a comparison of Cochrane reviews with articles published in paper-based journals. *JAMA*. 1998;280:278–280.
8. Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. *Ann Intern Med*. 2001;135:982–989.
9. Moher D, Cook DJ, Eastwood S, et al. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Onkologie*. 2000;23:597–602.
10. Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet*. 1998;352:609–613.
11. Schulz KF, Chalmers I, Hayes RJ, et al. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA*. 1995;273:408–412.
12. Sterne JA, Juni P, Schulz KF, et al. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. *Stat Med*. 2002;21:1513–1524.
13. Verhagen AP, de Vet HC, de Bie RA, et al. The art of quality assessment of RCTs included in systematic reviews. *J Clin Epidemiol*. 2001;54:651–654.
14. Egger M, Davey Smith G, Schneider M, et al. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997;315:629–634.
15. Furukawa TA, Streiner DL, Hori S. Discrepancies among megatrials. *J Clin Epidemiol*. 2000;53:1193–1199.
16. Villar J, Carroli G, Belizan JM. Predictive ability of meta-analyses of randomised controlled trials. *Lancet*. 1995;345:772–776.
17. LeLorier J, Gogoi G, Benhaddad A, et al. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *N Engl J Med*. 1997;337:536–542.
18. Sacks HS, Berrier J, Reitman D, et al. Meta-analyses of randomized controlled trials. *N Engl J Med*. 1987;316:450–455.
19. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. *J Clin Epidemiol*. 1991;44:1271–1278.
20. Repair of groin hernia with synthetic mesh: meta-analysis of randomized controlled trials. *Ann Surg*. 2002;235:322–332.
21. Oxman AD, Guyatt GH, Singer J, et al. Agreement among reviewers of review articles. *J Clin Epidemiol*. 1991;44:91–98.
22. Moher D, Jadad AR, Tugwell P. Assessing the quality of randomized controlled trials: current issues and future directions. *Int J Technol Assess Health Care*. 1996;12:195–208.
23. Smith GD, Altman DG, Egger M. *Systematic Reviews in Health Care: Meta-Analysis in Context*. 2nd ed. London: BMJ; 2001.
24. Clarke M. The QUORUM statement. *Lancet*. 2000;355:756–757.
25. Chalmers TC, Frank CS, Reitman D. Minimizing the three stages of publication bias. *JAMA*. 1990;263:1392–1395.
26. Kelly KD, Travers A, Dorgan M, et al. Evaluating the quality of systematic reviews in the emergency medicine literature. *Ann Emerg Med*. 2001;38:518–526.
27. Barnes J, Resch KL, Ernst E. Homeopathy for postoperative ileus? a meta-analysis. *J Clin Gastroenterol*. 1997;25:628–633.
28. Irwig L, Bennetts A. Quality of life after breast conservation or mastectomy: a systematic review. *Aust N Z J Surg*. 1997;67:750–754.
29. Koch A, Bouges S, Ziegler S, et al. Low molecular weight heparin and unfractionated heparin in thrombosis prophylaxis after major surgical intervention: update of previous meta-analyses. *Br J Surg*. 1997;84:750–759.
30. Li H, Yao SC. Surgical treatment for carcinoma of the oesophagus in Chinese language publications. *Br J Surg*. 1997;84:855–857.
31. Morris AD, Morris RD, Wilson JF, et al. Breast-conserving therapy vs mastectomy in early-stage breast cancer: a meta-analysis of 10-year survival. *Cancer J Sci Am*. 1997;3:6–12.
32. Munnangi S, Sonnenberg A. Colorectal cancer after gastric surgery: a meta-analysis. *Am J Gastroenterol*. 1997;92:109–113.
33. Palmer AJ, Schramm W, Kirchhof B, et al. Low molecular weight heparin and unfractionated heparin for prevention of thrombo-embolism in general surgery: a meta-analysis of randomised clinical trials. *Hae-mostasis*. 1997;27:65–74.
34. Cheek CM, Black NA, Devlin HB, et al. Groin hernia surgery: a systematic review. *Ann R Coll Surg Engl*. 1998;80(suppl 1):S1–80.
35. MacRae HM, McLeod RS. Handsewn vs. stapled anastomoses in colon and rectal surgery: a meta-analysis. *Dis Colon Rectum*. 1998;41:180–189.
36. McAlister FA, Clark HD, Wells PS, et al. Perioperative allogeneic blood transfusion does not cause adverse sequelae in patients with cancer: a meta-analysis of unconfounded studies. *Br J Surg*. 1998;85:171–178.
37. Platell C, Hall J. What is the role of mechanical bowel preparation in patients undergoing colorectal surgery? *Dis Colon Rectum*. 1998;41:875–882.
38. Rosen M, Chan L, Beart RW Jr, et al. Follow-up of colorectal cancer: a meta-analysis. *Dis Colon Rectum*. 1998;41:1116–1126.
39. Vanek VW. Meta-analysis of effectiveness of intermittent pneumatic compression devices with a comparison of thigh-high to knee-high sleeves. *Am Surg*. 1998;64:1050–1058.
40. Weiland DE, Bay RC, Del Sordi S. Choosing the best abdominal closure by meta-analysis. *Am J Surg*. 1998;176:666–670.
41. Boyages J, Delaney G, Taylor R. Predictors of local recurrence after treatment of ductal carcinoma in situ: a meta-analysis. *Cancer*. 1999;85:616–628.
42. Burroughs AK, Patch D. Transjugular intrahepatic portosystemic shunt. *Semin Liver Dis*. 1999;19:457–473.
43. Chung RS, Rowland DY. Meta-analyses of randomized controlled trials of laparoscopic vs conventional inguinal hernia repairs. *Surg Endosc*. 1999;13:689–694.
44. Garbutt JM, Soper NJ, Shannon WD, et al. Meta-analysis of randomized controlled trials comparing laparoscopic and open appendectomy. *Surg Laparosc Endosc*. 1999;9:17–26.
45. Heys SD, Walker LG, Smith I, et al. Enteral nutritional supplementation with key nutrients in patients with critical illness and cancer: a meta-analysis of randomized controlled clinical trials. *Ann Surg*. 1999;229:467–477.
46. Nelson RL. Meta-analysis of operative techniques for fissure-in-ano. *Dis Colon Rectum*. 1999;42:1424–1428.
47. Orr RK. The impact of prophylactic axillary node dissection on breast cancer survival: a Bayesian meta-analysis. *Ann Surg Oncol*. 1999;6:109–116.
48. Rosenberg L, MacNeil P, Turcotte L. Economic evaluation of the use of octreotide for prevention of complications following pancreatic resection. *J Gastrointest Surg*. 1999;3:225–232.
49. Temple LK, Litwin DE, McLeod RS. A meta-analysis of laparoscopic versus open appendectomy in patients suspected of having acute appendicitis. *Can J Surg*. 1999;42:377–383.
50. Urbach DR, Kennedy ED, Cohen MM. Colon and rectal anastomoses do not require routine drainage: a systematic review and meta-analysis. *Ann Surg*. 1999;229:174–180.
51. Wiseman DM, Trout JR, Franklin RR, et al. Meta-analysis of the safety and efficacy of an adhesion barrier (Interceed TC7) in laparotomy. *J Reprod Med*. 1999;44:325–331.
52. Derossis AM, DaRosa DA, Dutta S, et al. A ten-year analysis of surgical education research. *Am J Surg*. 2000;180:58–61.
53. Hodgson NC, Malthaner RA, Ostbye T. The search for an ideal method of abdominal fascial closure: a meta-analysis. *Ann Surg*. 2000;231:436–442.
54. Miltenburg DM, Prost HM, Graviss EA, et al. The role of frozen section, gender, age, and tumor size in the differentiation of follicular adenoma from carcinoma: a meta-analysis. *Surgery*. 2000;128:1075–1081.
55. Monteforte MJ, Turkelson CM. Bariatric surgery for morbid obesity. *Obes Surg*. 2000;10:391–401.
56. Palit TK, Miller CC 3rd, Miltenburg DM. The efficacy of thyroidectomy for Graves' disease: a meta-analysis. *J Surg Res*. 2000;90:161–165.
57. Rucinski J, Fabian T, Panagopoulos G, et al. Gangrenous and perforated appendicitis: a meta-analytic study of 2532 patients indicates that the incision should be closed primarily. *Surgery*. 2000;127:136–141.
58. Sauerland S, Nagelschmidt M, Mallmann P, et al. Risks and benefits of preoperative high dose methylprednisolone in surgical patients: a systematic review. *Drug Saf*. 2000;23:449–461.
59. Stocchi L, Nelson H. Wound recurrences following laparoscopic-assisted colectomy for cancer. *Arch Surg*. 2000;135:948–958.
60. Taylor MC, McLeod RS, Langer B. Biliary stenting versus bypass surgery for the palliation of malignant distal bile duct obstruction: a meta-analysis. *Liver Transpl*. 2000;6:302–308.

61. Tichansky D, Cagir B, Yoo E, et al. Strictureplasty for Crohn's disease: meta-analysis. *Dis Colon Rectum*. 2000;43:911–919.
62. Verkooijen HM, Peeters PH, Buskens E, et al. Diagnostic accuracy of large-core needle biopsy for nonpalpable breast disease: a meta-analysis. *Br J Cancer*. 2000;82:1017–1021.
63. Chapman AE, Levitt MD, Hewett P, et al. Laparoscopic-assisted resection of colorectal malignancies: a systematic review. *Ann Surg*. 2001;234:590–606.
64. Hulscher JB, Tijssen JG, Obertop H, et al. Transthoracic versus transhiatal resection for carcinoma of the esophagus: a meta-analysis. *Ann Thorac Surg*. 2001;72:306–313.
65. Lewis SJ, Egger M, Sylvester PA, et al. Early enteral feeding versus “nil by mouth” after gastrointestinal surgery: systematic review and meta-analysis of controlled trials. *BMJ*. 2001;323:773–776.
66. Mismetti P, Laporte S, Darmon JY, et al. Meta-analysis of low molecular weight heparin in the prevention of venous thromboembolism in general surgery. *Br J Surg*. 2001;88:913–930.
67. Rucinski J, Margolis M, Panagopoulos G, et al. Closure of the abdominal midline fascia: meta-analysis delineates the optimal technique. *Am Surg*. 2001;67:421–426.
68. Urschel JD, Urschel DM, Miller JD, et al. A meta-analysis of randomized controlled trials of route of reconstruction after esophagectomy for cancer. *Am J Surg*. 2001;182:470–475.
69. Urschel JD, Blewett CJ, Bennett WF, et al. Handsewn or stapled esophagogastric anastomoses after esophagectomy for cancer: meta-analysis of randomized controlled trials. *Dis Esophagus*. 2001;14:212–217.
70. Carless PA, Anthony DM, Henry DA. Systematic review of the use of fibrin sealant to minimize perioperative allogeneic blood transfusion. *Br J Surg*. 2002;89:695–703.
71. Lens MB, Dawes M, Goodacre T, et al. Elective lymph node dissection in patients with melanoma: systematic review and meta-analysis of randomized controlled trials. *Arch Surg*. 2002;137:458–461.
72. Lewis RT. Oral versus systemic antibiotic prophylaxis in elective colon surgery: a randomized study and meta-analysis send a message from the 1990s. *Can J Surg*. 2002;45:173–180.
73. Ludwig K, Bernhardt J, Steffen H, et al. Contribution of intraoperative cholangiography to incidence and outcome of common bile duct injuries during laparoscopic cholecystectomy. *Surg Endosc*. 2002;16:1098–1104.
74. Renehan AG, Egger M, Saunders MP, et al. Impact on survival of intensive follow up after curative resection for colorectal cancer: systematic review and meta-analysis of randomised trials. *BMJ*. 2002;324:813.
75. Sewnath ME, Karsten TM, Prins MH, et al. A meta-analysis on the efficacy of preoperative biliary drainage for tumors causing obstructive jaundice. *Ann Surg*. 2002;236:17–27.
76. Voyles CR, Hamilton BJ, Johnson WD, et al. Meta-analysis of laparoscopic inguinal hernia trials favors open hernia repair with preperitoneal mesh prosthesis. *Am J Surg*. 2002;184:6–10.